

ABSTRAK

Perkembangan pesat aplikasi ChatGPT di Indonesia mendorong meningkatnya jumlah ulasan pengguna di Google Play Store yang berpotensi menjadi sumber informasi berharga untuk memahami persepsi publik. Namun, besarnya volume ulasan dan ketidakseimbangan data antara sentimen positif dan negatif menjadi tantangan dalam proses analisis secara otomatis. Penelitian ini bertujuan untuk menganalisis kinerja algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan aplikasi ChatGPT serta mengetahui jumlah topik optimal dan nilai *coherence score* dari pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) pada ulasan positif dan negatif. Dataset yang digunakan terdiri dari 5.000 ulasan berbahasa Indonesia yang diperoleh dari Google Play Store. Hasil pelabelan manual menghasilkan 3.815 ulasan positif (76,3%) dan 1.185 ulasan negatif (23,7%). Validasi dilakukan oleh pakar bahasa untuk memastikan ketepatan pelabelan. Data terlebih dahulu melalui tahap *preprocessing* meliputi *cleaning*, *case folding*, tokenisasi, normalisasi, *stopword removal*, dan *stemming*. Klasifikasi sentimen dilakukan dalam empat skenario pengujian menggunakan SVM dengan kernel RBF, yaitu tanpa SMOTE, dengan SMOTE, tanpa SMOTE dengan optimasi parameter, dan dengan SMOTE dan optimasi parameter. Pemodelan topik menggunakan LDA dilakukan secara terpisah pada ulasan positif dan negatif dengan penentuan jumlah topik optimal berdasarkan *coherence score* CV. Hasil penelitian menunjukkan bahwa skenario terbaik pada klasifikasi SVM diperoleh dari kombinasi SMOTE dan optimasi parameter *grid search* ($C=10$, $\gamma=0,01$) dengan akurasi 84,28%, *precision* 86%, *recall* 84%, dan *F1-score* 85%. Pada pemodelan topik LDA, ulasan positif menghasilkan 10 topik optimal dengan *coherence score* 0,571 sedangkan ulasan negatif menghasilkan 2 topik optimal dengan *coherence score* 0,524. Hasil ini menunjukkan bahwa kombinasi SVM dan LDA mampu memberikan analisis yang menyeluruh terhadap sentimen dan topik ulasan pengguna aplikasi ChatGPT.

Kata kunci: Analisis Sentimen, *Support Vector Machine*, SMOTE, *Latent Dirichlet Allocation*, *Text Mining*

ABSTRACT

The rapid development of the ChatGPT application in Indonesia has driven an increasing number of user reviews on the Google Play Store, which potentially serve as a valuable source of information for understanding public perception. However, the large volume of reviews and data imbalance between positive and negative sentiments pose challenges in the automatic analysis process. This study aims to analyze the performance of the Support Vector Machine (SVM) algorithm in classifying the sentiment of ChatGPT application reviews, as well as to determine the optimal number of topics and coherence score values from topic modeling using Latent Dirichlet Allocation (LDA) on positive and negative reviews. The dataset consists of 5,000 Indonesian-language reviews obtained from the Google Play Store. Manual labeling yielded 3,815 positive reviews (76.3%) and 1,185 negative reviews (23.7%). Validation was conducted by a language expert to ensure labeling accuracy. The data first underwent a preprocessing stage comprising cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Sentiment classification was carried out in four testing scenarios using SVM with an RBF kernel, namely without SMOTE, with SMOTE, without SMOTE with parameter optimization, and with SMOTE and parameter optimization. Topic modeling using LDA was performed separately on positive and negative reviews, with the optimal number of topics determined based on the CV coherence score. The results show that the best scenario in SVM classification was obtained from the combination of SMOTE and grid search parameter optimization ($C=10$, $\gamma=0.01$), achieving an accuracy of 84.28%, precision of 86%, recall of 84%, and F1-score of 85%. In LDA topic modeling, positive reviews yielded 10 optimal topics with a coherence score of 0.571, while negative reviews yielded 2 optimal topics with a coherence score of 0.524. These results indicate that the combination of SVM and LDA is capable of providing a comprehensive analysis of the sentiment and topics of ChatGPT application user reviews..

Keywords: Sentiment Analysis, Support Vector Machine, SMOTE, Latent Dirichlet Allocation, Text Mining